Nonlinearities in the development process: A nonparametric approach

Théophile T. Azomahou
UNU-MERIT, Maastricht University

Globelics Academy 2008
Tampere, Finland

Parametric Vs. nonparametric models

Accounting for nonlinearities in parametric and nonparametric regressions

Parametric linear regression model

$$y = \alpha + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \varepsilon$$

- Aim: estimate the parameters α and βs.
- Advantage: basic econometrics, easy to perform, friendly user software available (methods: least squares, maximum likelihood, IV, GMM, ...)
- Drawback: misspecification problem (DGP is assumed to be known).

Parametric Vs. nonparametric models

Accounting for nonlinearities in parametric and nonparametric regressions

Nonparametric (or semiparametric) regression

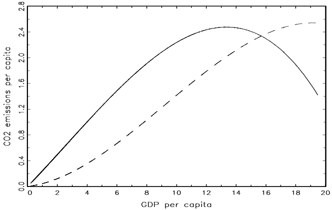
$$y = f(x) + \varepsilon$$

- Aim: estimate the function $f(x)$, without assuming a particular form for $f(x)$.
- Advantage: robust to misspecification, DGP unknown (methods: kernel, cubic spline, k-nn, series, local linear, etc).
- Drawback: possibly data consuming, lack of friendly user procedures.

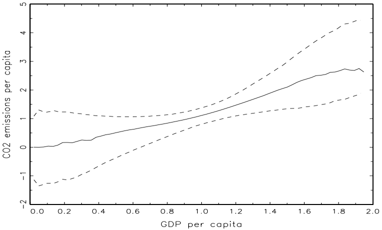
Examples (cont'd)

1. The case of CO_2 emissions (EKC)

- Parametric estimation:
  inverted U-shaped relationship



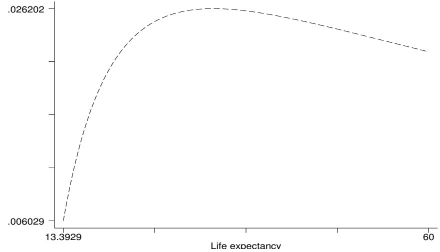
- Nonparametric estimation:
  monotonically increasing function



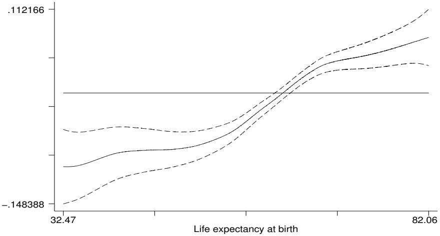
Examples

2. Life expectancy and income growth

- Suggested parametric: hump-shaped relationship between life expectancy and growth.



- Nonparametric: convex-concave relationship between life expectancy at birth and income per capita.



Outline of the presentation

1. Set up
2. Kernel density estimation
3. Nonparametric regression
4. Semiparametric estimation
5. Application: technology frontier

Kernel density estimation

The core of the method: The kernel density estimation

Density as distribution

- The density of a variable describes the distribution of the values that the random variable takes.
  - Fully parametric distribution assumes about the form of the density.
  - Canonical example: if $X \sim N(\mu, \sigma^2)$, then

$$\hat{f}(x) = f(x \,|\, \hat{\mu}, \hat{\sigma}^2) = \frac{1}{\hat{\sigma}} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\hat{\mu}}{\hat{\sigma}}\right)^2\right].$$

    for some estimation of $\hat{\mu}$ and $\hat{\sigma}$ (mean and variance obtained from a given sample).
- Problem: narrow distributional assumption about the density.

Kernel density estimation

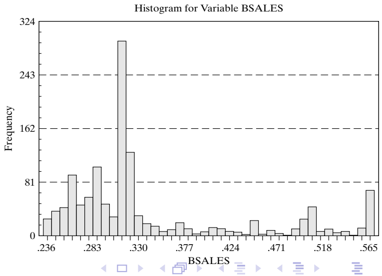
Histogram as a crude density estimator

Example of distribution of sales over 1,270 firms

- Descriptive statistics:

Variables	Mean	Std.Dev.	Min.	Max.
Sales	0.3428	0.08919	0.2361	0.5664

- Histogram:
  The distribution seems to be **bimodal**, but no particular functional form seems natural



Kernel density estimation

From histogram to kernel

Frequency

$$\hat{f}(x) = \frac{1}{n}\frac{\text{frequency in bin}_x}{\text{width of bin}_x} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}\mathbf{1}\left(x - \frac{h}{2} < x_i < x + \frac{h}{2}\right)$$

where:

- x_k is the midpoint of the kth bin and h is the width of the bin. The distance to the left and right boundaries of the bins are $h/2$.
- $\mathbf{1}(\textit{statement})$ denotes an indicator function.
- The frequency count in each bin is the number of observations in the sample which fall in the range $x_k \pm h/2$. Collecting terms gives the formula.
- bin_x denotes the bin which has x as its midpoint.

Kernel density estimation

From histogram to kernel

Rearrange the event in the indicator function to produce an equivalent form: the **(naive) density kernel estimator**.

$$\begin{aligned} \hat{f}(x) &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}\mathbf{1}\left(-\frac{1}{2} < \frac{x_i - x}{h} < \frac{1}{2}\right) \\ &= \hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left[\frac{x_i - x}{h}\right] \end{aligned}$$

where:

- $K[z] = \mathbf{1}[-1/2 < z < 1/2]$.
- This form of the estimator counts the number of points that are within $^1/_2$ bin width of x_k.

Kernel density estimation

i) From histogram to kernel: why is it naive?

- This estimator is neither smooth nor continuous (crudeness of $K[z]$).
- Its shape is partly determined by where the leftmost and rightmost terminals of the histogram are set.
- The shape of the histogram will be crucially dependent on the bandwidth, itself.

ii) How to overcome the crudeness of the weighting function $K[z]$?

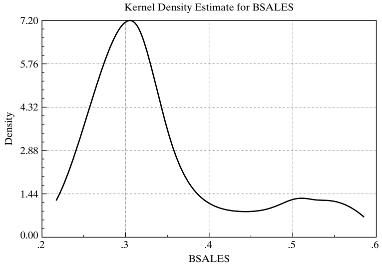
- Rosenblatt (1956): substitute for the naive estimator some other weighting function which is continuous and integrates to one.

Kernels	**Formula K[z]**
Epanechnikov	$0,75(1-.2z^2)/2,236$ if $\lvert z\rvert \leq 5,\ 0$ elsewhere
Normal	$\phi(z)$ (normal density),
Uniform	0,5 if $\lvert z\rvert \leq 1,\ 0$ elsewhere
Logit	$\Lambda(z)[1-\Lambda(z)]$ (logistic density)
Parzen	$4/3-8z^2+8\lvert z\rvert^3$ if $\lvert z\rvert \leq 0,5,\ 8(1-\lvert z\rvert)^3/3$ elsewhere

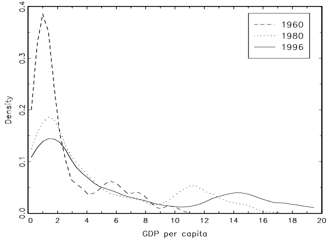
Kernel density estimation: Back to data (applications)

.

- Distribution of sales:
  Logistic kernel



- Distribution of GDP per cap:
  Epanechnikov kernel



Nonparametric regression

1. Regression function of a variable y on a single variable x:

$$y = m(x) + \varepsilon$$

   - No assumptions about distribution, homoscedasticity, serial correlation.
   - The functional form is still the same for all values of x but unknown.

2. Methods: smoothing techniques
   - **The case of Nadaraya-Waston estimator**

$$\hat{m}(x^* \mid \mathbf{x}, h) = \frac{\sum_{i=1}^{n} \frac{1}{h} K\left[\frac{x_i - x^*}{h}\right] y_i}{\sum_{i=1}^{n} \frac{1}{h} K\left[\frac{x_i - x^*}{h}\right] = \hat{f}(x)},$$

Nonparametric regression

1. Easy to implement: A GAUSS procedure (only two lines!)

```
proc (2) = npr(y,x);
  local reg, i, f;
  i=1;
  reg=zeros(n,1);
  f=zeros(n,1);
  do until i>n;
    f[i,1]=sumc(pdfn((x-x[i,1])/h))/(n*h);
    reg[i,1]=(sumc(pdfn((x-x[i,1])/h).*y)/(n*h))/f[i,1];
  i=i+1;
  endo;
  retp(f,reg);
endp;
```

2. Shortcomings
    - Curse of dimensionality
    - Slow speed of convergence
    - Possibly data consuming

Semiparametric estimation: The partially linear regression

1. Consider the specification:

$$y = f(x) + \mathbf{z}'\boldsymbol{\beta} + \varepsilon,$$

   Take a modified version of the previous:

$$y - E(y|x) = [\mathbf{z} - E(\mathbf{z}|x)]\boldsymbol{\beta} + [\varepsilon - E(\varepsilon|x)]$$

2. Estimation procedure (Robinson, 1988):
   - **Step 1**: Compute nonparametric estimators for $E(y|x)$ and $E(\mathbf{z}|x)$ using the kernel method.
   - **Step 2**: Compute an estimator for $\boldsymbol{\beta}$, $\widehat{\boldsymbol{\beta}}$, by regressing $y - E(y|x)$ on $\mathbf{z} - E(\mathbf{z}|x)$. This step may be done by OLS.
   - **Step 3**: Finally, obtain an estimator of $f(x)$, $\widehat{f}(x)$, by a nonparametric regression $E\left[\left(y - \mathbf{z}'\widehat{\beta}\right)|x\right]$.

Application: technology frontier, labor productivity and economic growth

1. Objective
   - Study the growth strategy when countries are close to the technology frontier (US as reference).
   - Estimation of two models: In the first model, the dependent variable is GDP growth rate per worker (as a measure of labor productivity growth), and in the second, the dependent variable is labor productivity backwardness (in logarithmic term).
2. Data
   - 29 OECD countries data over the period 1960-2000.
   - Sources: Penn World Table 6.1, World Development Indicators and Eurostat.

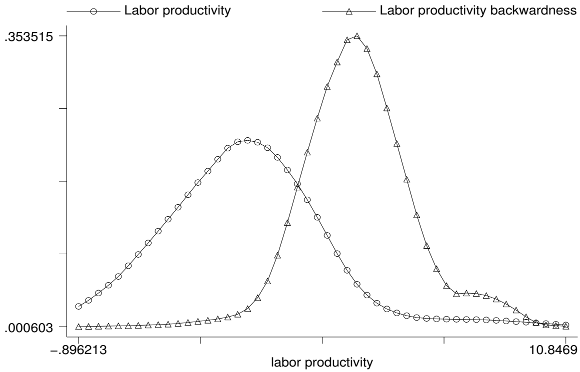
Application: technology frontier, labor productivity and economic growth

3. Descriptive statistics

Variables	#Obs.	Mean	Std.Dev.	Min.	Max.
Labor produc. growth rate	958	0.036	0.47	-0.01	1.02
Labor produc. backwardness	985	-0.88	1.2	-6.82	3.38
Primary school enrol. rate	603	0.11	0.01	0.08	0.14
Secondary school enrol. rate	615	0.85	0.23	0.11	1.48
School enrol. rate in higher educ.	280	0.36	0.19	0.04	0.98
Government R&D expenditure	196	0.4	0.14	0	0.7
Industry R&D expenditure	186	0.5	0.15	0	0.91
R&D expenditure from abroad	182	0.61	0.52	0	3.03

Application: technology frontier, labor productivity and economic growth

4. Distribution of variables of interest (kernel density estimation)



Application: technology frontier, labor produc. and growth

Specification: The Generalized Additive Model (GAM) for panel data

$$Y = \alpha + \sum_{j=1}^{p} f_j\left(X_j\right) + \mathbf{Z}'\gamma + \epsilon$$

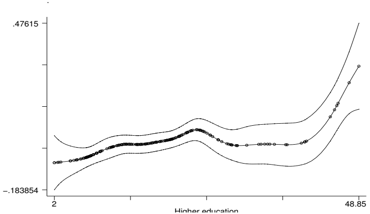
where:

- $Y = (y_{i1}, \cdots, y_{iT})'$ denotes the response variable
- $X_j = (x_{i1}, \cdots, x_{iT})'$ for $j = 1, \cdots, p$ are non linear explanatory variables, $i = 1, \cdots, n$ and $t = 1, \cdots, T$
- $\mathbf{Z}$ is the row vector of parametric components
- α denotes the regression intercept, and γ the vector of parameters
- The f_j are unknown univariate functions to be estimated such that $\mathbb{E}\left[f_j\left(Xj\right)\right] = 0$.
- error: $\epsilon = (\epsilon_{i1}, \cdots, \epsilon_{iT})'$ is such that $\mathbb{E}(\epsilon|X_1, ..., X_p, \mathbf{Z}) = 0$ and $\mathbb{V}(\epsilon|X_1, ..., X_p, \mathbf{Z}) = \sigma^2(X_j, \mathbf{Z})$

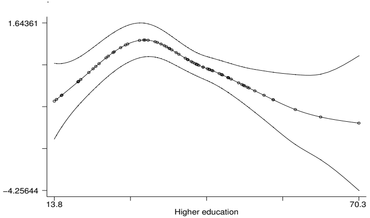
Application: technology frontier, labor productivity and economic growth

5. Estimation results (cont'd)

- Labor productivity growth and school enrollment rate in higher education



- Labor productivity backwardness and school enrollment rate in higher education



Application: technology frontier, labor produc. and growth

5. Estimation results

Labor productivity backwardness and the part of R&D expenditure in % of GERD funded by industries

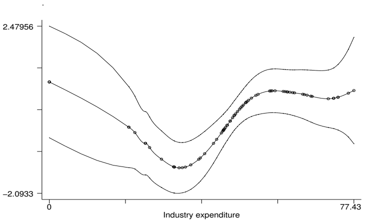


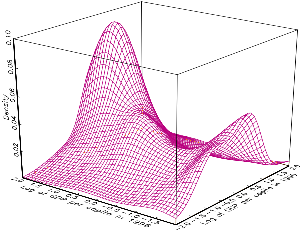
Table: Semipamatric estimation for labor productivity backwardness

Variables	Coef.	Std.Err	df.	Gain[(a)]
Primary school enrollment rate	-0.08*	0.04	2	0.64
Secondary school enrollment rate	-0.02	0.013	2	1.8
School enrollment rate in higher education	-0.02	0.02	4.99	20.29[(b)]
Government R&D expenditure	0.03**	0.01	3.99	1.24
Industry R&D expenditure	0.02	0.011	5.99	20.39[(b)]
R&D expenditure from abroad	-0.04	0.03	2	0.17
Intercept	-0.18	0.16	1	-

Regional convergence

The dynamics of transition of regional GDP in Europe

Conditionnal density of GDP per capita



Polarization of GDP per capita